Peter T. Chernek

Review Section 3.3

Sections 3.4-3.6

Causal Inference: What if

Miguel A. Hernán, James M. Robins

Section 3.3- Positivity

- **Positivity Condition-** We must ensure that there is a probability greater than zero–a positive probability–of being assigned to each of the treatment levels A=1 and A=0.
- **Alternative-** The investigator assigns all individuals to either A = 1 or A= 0. How can we compute any type of causal effect?
- **Marginally Randomized Studies-** Both pr(A=1) and pr(A=0) are positive by design
- **Conditionally Randomized Studies-**Pr [A = 1|L= *l*] and Pr [A = 0|L = *l*] are also positive by design

Conditionally Randomized Study and Positivity

- If **Pr[A=*a*|L = *l*] > 0** for all *a* involved in the causal contrast then we have positivity in our study
- We can verify all 4 probabilities since there is only one binary covariate L
- **Cases *a*=1 :** Pr [A = 1|L = 1] = 0.75 and Pr [A = 1|L = 0] = 0.50
- **Cases *a*=0 :** Pr [A = 0|L = 0] = 0.50 and Pr [A = 0|L = 1]=0.25
- Only required for values of *L* in the population of interest, in this case just L=0 and L=1
- Also, only required for variables that predict **Y** or used to assign treatment, not "**whether the person has blue eyes**" for example since this is not associated with the treatment or outcome.

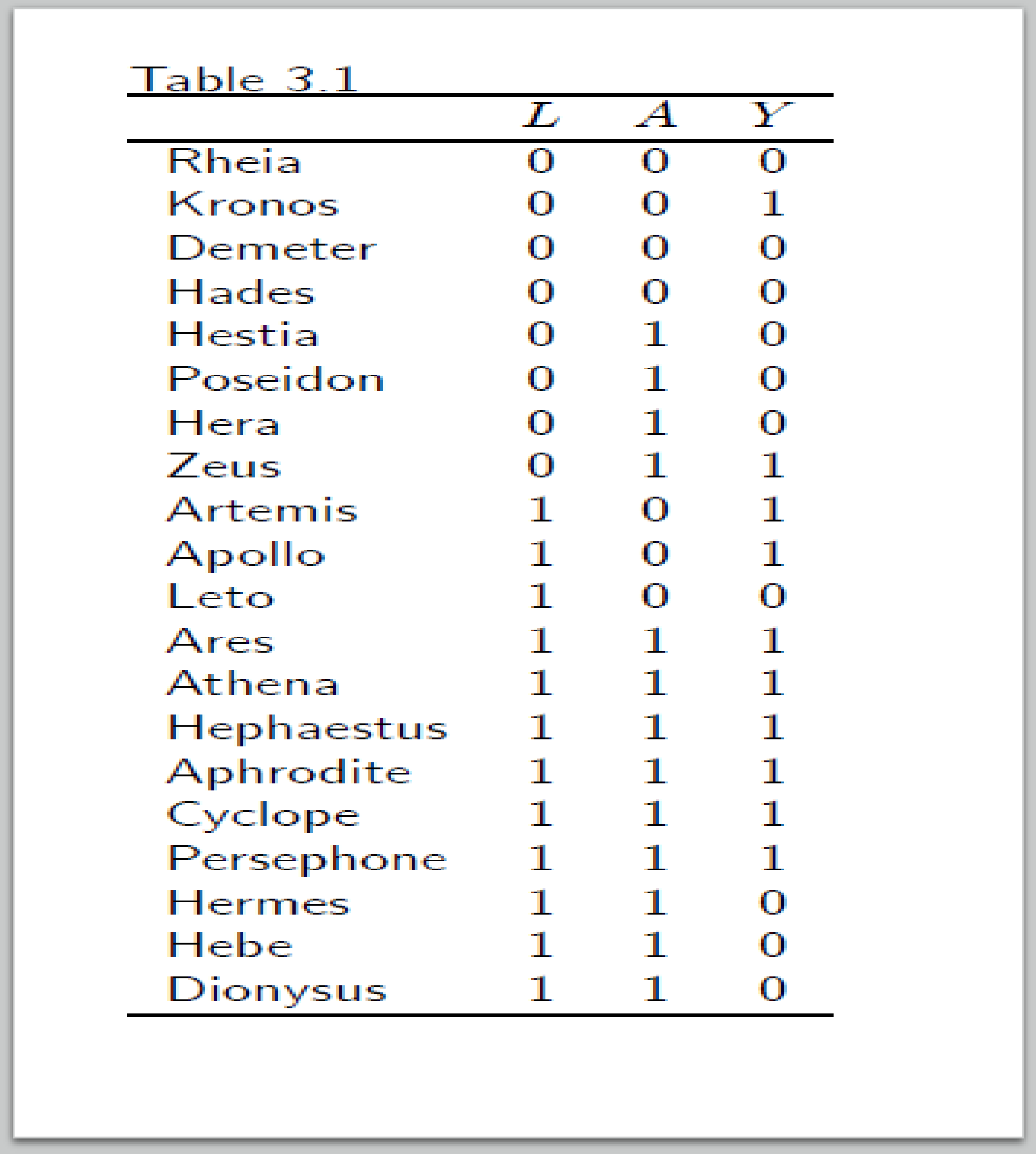
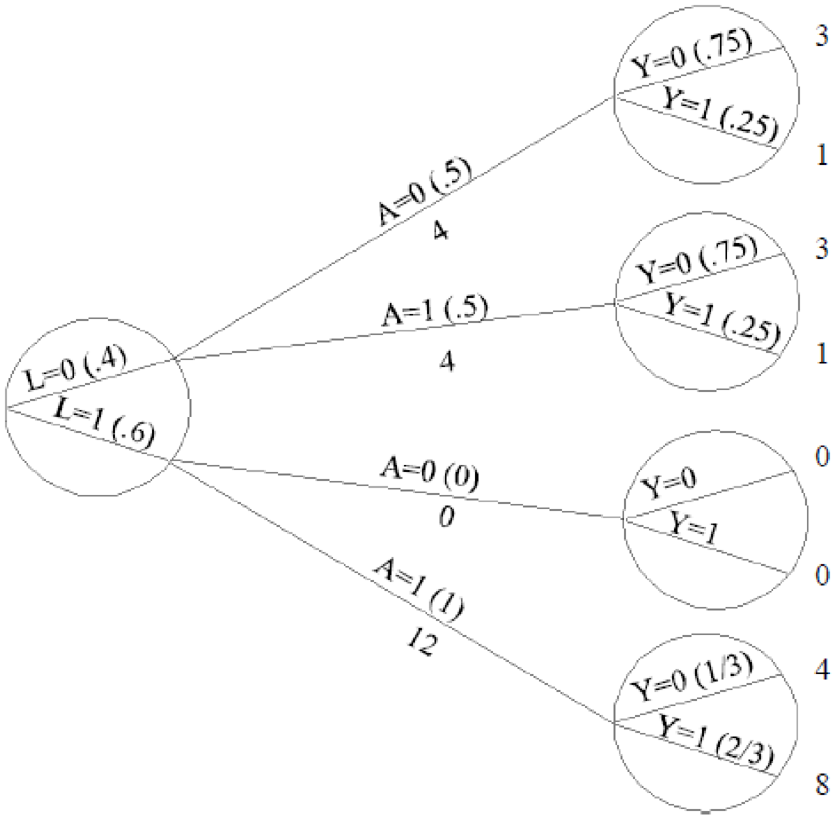


Table 3.1

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

- When there is **no randomization** as in an observational study positivity and exchangeability are never guaranteed
- For example, doctors could have transplanted a heart only to those in critical condition and therefore **Pr [A = 0|L = 1] = 0**
- Previously, IP weighting and Standardization only assumed exchangeability
- These concepts are only meaningful when positivity holds as well, why?
- We have no untreated individuals in **L=1** to show what would have happened to the 12 people who were treated had they not been treated.



Consistency
Section 3.4

- [**Y*a* = Y]** for every individual with **A = *a***
- ***a*=1 :** The **observed** outcome for every **treated** individual equals their outcome **if** they had received treatment
- ***a*=0** : The **observed** outcome for every **untreated** individual equals their outcome **if** they had remained untreated

Defining Treatment *A*

- How we define the treatment is very important to causal inference
- Suppose that we want to quantify the causal effect of **Obesity A** at age 40 on the risk of **Mortality Y** by age 50 in a certain population
- There are multiple versions of the treatment A = 1 defined by duration, recency, and intensity of obesity
- What if a person was obese for 20 years prior to age 40 versus obese for 2 years prior to age 40?
- Each of these versions may have a different effect on mortality at age 50, so we need to provide a detailed definition of the version of obesity at age 40 that we are using in the study
- The causal effect **Pr[Y*a*=1 = 1] − Pr[Y*a*=0 =1]** will not be well defined if the treatment is not well defined

Defining a Multi-Factor Treatment *A*

- **Case 1**: Zeus had **genes that predisposed him to large amounts of fat tissue**, but exercised moderately, keeps a healthy diet, and has a favorable intestinal microbiota. Considered obese at age 40 and dies at age 49

- **Case 2:** Zeus has neutral genes, but has a **lifetime of lack of exercise**, **too many calories in the diet**, and an **unfavorable intestinal microbiota**. Considered obese at 40 but survives past age 50

- Without a very precise definition of obesity or ***a=1,*** the value of ***Ya*** at age 50 can differ depending on the factor related to the obesity at age 40.

Sufficiently Well-Defined Interventions

- Variation in treatment is acceptable as long as it does not alter the value of **Y*a***. (i.e. Running around a park in different directions should not alter **Y*a,*** although these are different treatments per se.)
- Declaring a treatment sufficiently well-defined is a matter of agreement among experts based on the available knowledge.
- Future research can alter what today is considered not to matter in effecting the outcome (i.e. moving the the body to the right, but not to the left, while running may be harmful and so the direction of running could have different effects on the outcome **Y*a***)

Linking Counterfactuals to The Observed Data

- **A=1:** At age 18 and through age 40, put every individual on a stringent mandatory diet that guarantees that they would never weigh more than their weight at the age of 18 years. Whenever the weight is greater than the baseline weight at 18 years, the individual's caloric intake is restricted, without changing his usual mix of calorie sources and micronutrients, until the time (usually within 1—3 days) that the individual falls below baseline weight

- **A=0 :** The comparison intervention is "do not intervene"

Does Y*a*=Y Hold for Individuals in this Study?

- Treatment values A = 1 and A = 0 are sufficiently well-defined and, therefore, no meaningful “vagueness” should remain in the specification of the counterfactual outcomes Y*a*=1 and Y*a*=0
- Take Ares, in the study population, who did not receive caloric treatment A=1 and who still maintained a constant weight between 18-40.
- Ares observed outcome **Y** does not necessarily equal is outcome **Y*a*** had he received caloric treatment
- Although the intervention is well defined, it needs to be linked to the observed data, that is the equality **Y*a* = Y** must hold for some individuals like Ares

Fine Point 3.3

Possible worlds. Some philosophers of science define causal contrasts using the concept of "possible worlds." The actual world is the way things actually are. A possible world is a way things might be. Imagine a possible world a where everybody receives treatment value a, and a possible world a' where everybody receives treatment value a'. The mean of the outcome is $\mathrm{E}[Y^a]$ in the first possible world and $\mathrm{E}[Y^{a'}]$ in the second one. These philosophers say that there is an average causal effect if $\mathrm{E}[Y^a] \neq \mathrm{E}[Y^{a'}]$ and the worlds a and a' are the two worlds closest to the actual world where all individuals receive treatment value a and a', respectively.

We introduced an individual's counterfactual outcome Y^a as her outcome under a sufficiently well-defined intervention that assigned treatment value a to her. These philosophers prefer to think of the counterfactual Y^a as the outcome in the possible world that is closest to our world and where the individual was treated with a. Both definitions are equivalent when the only difference between the closest possible world and the actual world is that the intervention of interest took place. The possible worlds formulation of counterfactuals replaces the sometimes difficult problem of specifying the intervention of interest by the equally difficult problem of describing the closest possible world that is minimally different from the actual world. Stalnaker (1968) and Lewis (1973) proposed counterfactual theories based on possible worlds.

- What happens if the data we collect for this study is only on body weight at age 40, but not on the individual’s lifetime history of weight, exercise, and diet?
- Restriction to the treatment value ***a*** of interest may be impossible when if our data are not “sufficiently rich” and we cannot get access to lifetime weight data
- One way out of this problem is to **assume** that the effects of **all** versions of treatment are identical–that is, if there is **Treatment-Variation irrelevance**
- May be a good approximation
- Lowering blood pressure through different pharmacological mechanisms results in similar outcomes therefore we may argue that a precise definition of the treatment is unnecessary to link the potential and observed outcomes.

Technical Point 3.2

Cheating consistency. Consider a compound treatment R with multiple, relevant versions of treatment. Interestingly, even if the versions of treatment are not well defined, we may still articulate a consistency condition that is guaranteed to hold (Hernán and VanderWeele, 2011; VanderWeele and Hernán, 2013): For individuals with $R_i = r$ we let $A_i(r)$ denote the version of treatment $R_i = r$ actually received by individual i; for individuals with $R_i \neq r$ we define $A_i(r) = 0$ so that $A_i(r) \in \{0\} \cup \mathcal{A}(r)$. The consistency condition then requires for all i,

$$Y_i = Y_i^{r,a(r)} \text{ when } R_i = r \text{ and } A_i(r) = a(r).$$

That is, the outcome for every individual who received a particular version of treatment $R = r$ equals his outcome if he had received that particular version of treatment. This statement is true by definition of version of treatment if we in fact define the counterfactual $Y_i^{r,a(r)}$ for individual i with $R_i = r$ and $A_i(r) = a(r)$ as individual i's outcome that he actually had under actual treatment r and actual version $a(r)$. However, using this consistency condition is self-defeating because, as discussed in the main text, it prevents us from understanding what effect is being estimated and from being able to evaluate exchangeability and positivity.

Similarly, consider the following hypothetical intervention: 'assign everybody to being nonobese by changing the determinants of body weight to reflect the distribution of those determinants in those who are nonobese in the study population.' This intervention would randomly assign a version of treatment to each individual in the study population so that the resulting distribution of versions of treatment exactly matches the distribution of versions of treatment in the study population. Analogously, we can propose another hypothetical, random intervention that assigns everybody to being obese.

This trick is implicitly used in the analysis of many observational studies that compare the risks $\Pr[Y = 1|A = 1]$ and $\Pr[Y = 1|A = 0]$ (often conditional on other variables) to endow the contrast with a causal interpretation. A problem with this trick is, of course, that the proposed random interventions may not match any realistic interventions we are interested in. Learning that intervening on 'the determinants of body weight to reflect the distribution of those determinants in those with nonobese weight' decreases mortality by, say, 30% does not imply that realistic interventions (e.g., modifying caloric intake or exercise levels) will decrease mortality by 30% too. In fact, if intervening on 'determinants of body weight in the population' requires intervening on genetic factors, then a 30% reduction in mortality may be unattainable by interventions that can actually be implemented in the real world.

- Need to carefully evaluate the version of treatment in population to detect mismatches between treatment values of interest and the data at hand.

- Not necessary if experts agree that all versions of the treatment in the data have the same causal effect.

- In summary, ill-defined treatments like “obesity” complicate the interpretation of causal effect estimates, but so do sufficiently well defined treatments that are absent in the data.

Section 3.6 Target Trial

- Causal effect refers to the contrast between average counterfactual outcomes under different treatment values
- Should be able to imagine a **hypothetical** randomized experiment to quantify this effect which is known as the **target trial**
- Observational data may be used to emulate such a trial when it can not be done for ethical or feasibility reasons
- **Step 1:** Describe such a trial
- **Step 2:** Show how the observational data can be used to emulate this trial

Example

- Consider the causal effect of “weight loss” on mortality in individuals who are obese and do not smoke at age 40
- An **explicit** emulation of the target trial would not be possible since a direct treatment such as a surgery to reduce BMI to certain level (25) is not feasible since few people undergo such a change in real world, and therefore it would be difficult to compare counterfactual outcomes to observed outcomes **How do we emulate this?**
- Estimate the effect of losing 5% of BMI every year, starting at age 40 and for as long as their BMI stays over 25, under the assumption that it does not matter how the weight loss is achieved
- Transfer this treatment strategy to the protocol of a target trial and use the observational data they have at hand to simulate the trial

- Do we need to specify a target trial or is just knowing that “deaths can be prevented from becoming non-obese no matter how” adequate enough?
- Authors claim if we do not specify the trial then this can lead to the lack of positivity and exchangeability.
- **How we lose out on exchangeability?** If we do not characterize the treatment version corresponding to our causal question about obesity then it is difficult to identify the covariates L that make obese and non-obese individuals conditionally exchangeable.
- **How does this effect positivity?** We may adjust for covariates L where some individuals will not be obese leading to the problem where there are no individuals receiving treatment in say L=1

- Analytic methods described in this book yield effect estimates that are only as well defined as the treatments that are being compared.

- Problems generated by unspecified treatments **cannot** be dealt with by applying sophisticated statistical methods according to authors.

- Even if target trial cannot be emulated then we can still focus on non-causal prediction although this may be unsatisfying to some.

- Obesity may predict or be associated with mortality the same way that carrying a lighter predicts lung cancer.

Attributable fraction. We have described effect measures like the causal risk ratio $\Pr[Y^{a=1}=1]/\Pr[Y^{a=0}=1]$ and the causal risk difference $\Pr[Y^{a=1}=1]-\Pr[Y^{a=0}=1]$. Both the causal risk ratio and the causal risk difference are examples of effect measures that compare the counterfactual risk under treatment $a=1$ with the counterfactual risk under treatment $a=0$. However, one could also be interested in measures that compare the observed risk with the counterfactual risk under either treatment $a=1$ or $a=0$. This contrast between observed and counterfactual risks allows us to compute the proportion of cases that are attributable to treatment in an observational study, i.e., the proportion of cases that would not have occurred had treatment not occurred. For example, suppose that all 20 individuals in our population attended a dinner in which they were served either ambrosia $(A=1)$ or nectar $(A=0)$. The following day, 7 of the 10 individuals who received $A=1$, and 1 of the 10 individuals who received $A=0$, were sick. For simplicity, assume exchangeability of the treated and the untreated so that the causal risk ratio is $0.7/0.1=7$ and the causal risk difference is $0.7-0.1=0.6$. (In conditionally randomized experiments, one would compute these effect measures via standardization or IP weighting.) It was later discovered that the ambrosia had been contaminated by a flock of doves, which explains the increased risk summarized by both the causal risk ratio and the causal risk difference. We now address the question 'what fraction of the cases was attributable to consuming ambrosia?'

In this study we observed 8 cases, i.e., the observed risk was $\Pr[Y=1]=8/20=0.4$. The risk that would have been observed if everybody had received $a=0$ is $\Pr[Y^{a=0}=1]=0.1$. The difference between these two risks is $0.4-0.1=0.3$. That is, there is an excess 30% of the individuals who did fall ill but would not have fallen ill if everybody in the population had received $a=0$ rather than their treatment A. Because $0.3/0.4=0.75$, we say that 75% of the cases are attributable to treatment $a=1$: compared with the 8 observed cases, only 2 cases would have occurred if everybody had received $a=0$. This *excess fraction* or *attributable fraction* is defined as

$$\frac{\Pr[Y=1]-\Pr[Y^{a=0}=1]}{\Pr[Y=1]}$$

Technical Point 3.1

Positivity for standardization and IP weighting. We have defined the standardized mean for treatment level a as $\sum_l \mathrm{E}\left[Y|A=a,L=l\right]\Pr\left[L=l\right]$. However, this expression can only be computed if the conditional quantity $\mathrm{E}\left[Y|A=a,L=l\right]$ is well defined, which will be the case when the conditional probability $\Pr\left[A=a|L=l\right]$ is greater than zero for all values l that occur in the population. That is, when positivity holds. (Note the statement $\Pr\left[A=a|L=l\right]>0$ for all l with $\Pr\left[L=l\right]\neq 0$ is effectively equivalent to $f\left[a|L\right]>0$ with probability 1.) Therefore, the standardized mean is defined as

$$\sum_l \mathrm{E}\left[Y|A=a,L=l\right]\Pr\left[L=l\right] \quad \text{if } \Pr\left[A=a|L=l\right]>0 \text{ for all } l \text{ with } \Pr\left[L=l\right]\neq 0,$$

and is undefined otherwise. The standardized mean can be computed only if, for each value of the covariate L in the population, there are some individuals that received the treatment level a.

The IP weighted mean $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[A|L\right]}\right]$ is no longer equal to $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[a|L\right]}\right]$ when positivity does not hold. Specifically, $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[a|L\right]}\right]$ is undefined because the undefined ratio $\frac{0}{0}$ occurs in computing the expectation. On the other hand, the IP weighted mean $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[A|L\right]}\right]$ is *always* well defined since its denominator $f\left[A|L\right]$ can never be zero. However, it is now a biased estimate of the counterfactual mean even under exchangeability. In particular, when positivity fails to hold, $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[A|L\right]}\right]$ is equal to $\Pr\left[L\in Q(a)\right]\sum_l \mathrm{E}\left[Y|A=a,L=l,L\in Q(a)\right]\Pr\left[L=l|L\in Q(a)\right]$ where $Q(a)=\{l;\Pr\left(A=a|L=l\right)>0\}$ is the set of values l for which $A=a$ may be observed with positive probability. Therefore, under exchangeability, $\mathrm{E}\left[\dfrac{I\left(A=a\right)Y}{f\left[A|L\right]}\right]$ equals $\mathrm{E}\left[Y^a|L\in Q(a)\right]\Pr\left[L\in Q(a)\right]$.

From the definition of $Q(a)$, $Q(0)$ cannot equal $Q(1)$ when A is binary and positivity does not hold. In this case the contrast $\mathrm{E}\left[\dfrac{I\left(A=1\right)Y}{f\left[A|L\right]}\right]-\mathrm{E}\left[\dfrac{I\left(A=0\right)Y}{f\left[A|L\right]}\right]$ has no causal interpretation, even under exchangeability, because it is a contrast between two different groups. Under positivity, $Q(1)=Q(0)$ and the contrast is the average causal effect if exchangeability holds.